Memory and visual search in naturalistic 2D and 3D environments

Chia-Ling Li The Institute for Neuroscience, The University of Texas at Austin, Austin, TX, USA

M. Pilar Aivar Facultad de Psicología, Universidad Autónoma de Madrid, Madrid, Spain

Dmitry M. Kit Department of Computer Science, University of Bath, Bath, UK

Matthew H. Tong Center for Perceptual Systems, The University of Texas at Austin, Austin, TX, USA

Mary M. Hayhoe Center for Perceptual Systems, The University of Texas at Austin, Austin, TX, USA

The role of memory in guiding attention allocation in daily behaviors is not well understood. In experiments with two-dimensional (2D) images, there is mixed evidence about the importance of memory. Because the stimulus context in laboratory experiments and daily behaviors differs extensively, we investigated the role of memory in visual search, in both two-dimensional (2D) and three-dimensional (3D) environments. A 3D immersive virtual apartment composed of two rooms was created, and a parallel 2D visual search experiment composed of snapshots from the 3D environment was developed. Eye movements were tracked in both experiments. Repeated searches for geometric objects were performed to assess the role of spatial memory. Subsequently, subjects searched for realistic context objects to test for incidental learning. Our results show that subjects learned the room-target associations in 3D but less so in 2D. Gaze was increasingly restricted to relevant regions of the room with experience in both settings. Search for local contextual objects, however, was not facilitated by early experience. Incidental fixations to context objects do not necessarily benefit search performance. Together, these results demonstrate that memory for global aspects of the environment guides search by restricting allocation of attention to likely regions, whereas task relevance determines what is learned from the active search experience. Behaviors in 2D and 3D environments are comparable, although there is greater use of memory in 3D.

Introduction

Allocation of attention is central to visual function in everyday behaviors. Normally, this process operates seamlessly so that critical information is attended at the appropriate time to control behavior. It seems likely that memory representations for familiar scenes are an important factor in normal attentional allocation. For example, it is not difficult to locate all the ingredients and tools you need to make yourself a meal in your own kitchen. In a friend's kitchen, however, it is likely to be more challenging because it is less familiar. Thus, memory representations of familiar environments may streamline the allocation of attention in everyday tasks and mitigate the effects of limited attentional resources.

To understand the role of memory in guiding attention, one focus of interest has been visual search. Memory representations encoding the relationship between objects in the scene have been shown to influence search (Castelhano & Heaven, 2011; Mack & Eckstein, 2011). When searching in a synthetic array of stimuli on a homogenous background (e.g., letters arranged in different orientations), search efficiency increases through implicit learning of the association between target and surrounding context, a form of learning termed *contextual cueing* (Chun & Jiang, 1998, 1999; Jiang & Wagner, 2004; Olson & Chun, 2002). Later studies have adopted images of naturalistic scenes

Citation: Li, C.-L., Aivar, M. P., Kit, D. M., Tong, M. H., & Hayhoe, M. M. (2016). Memory and visual search in naturalistic 2D and 3D environments. *Journal of Vision*, *16*(8):9, 1–20, doi:10.1167/16.8.9.

doi: 10.1167/16.8.9 Received January 31, 2016; published June 14, 2016 ISSN 1534-7362

as stimuli. In these cases, associations between scene and targets are learned much more rapidly (Brockmole, Castelhano, & Henderson, 2006; Brockmole & Henderson, 2006a, 2006b). When targets are realistic objects, one brief preview of the search scene is enough to facilitate search, even if search targets were absent during the preview (Castelhano & Henderson, 2007; Hollingworth, 2009; Võ & Henderson, 2010). When targets are embedded in the scenes during the preview, search performance further improves (Hollingworth, 2006, 2009), although the target effect was found to be small in another study (Castelhano & Henderson, 2007). The main effect of the preview is to guide the first two fixations to the relevant locations in the scene during search (Hillstrom, Scholey, Liversedge, & Benson, 2012). However, when information from scene semantics is available, search is primarily determined by this factor, and memory from previous exposures has little effect (Võ & Wolfe, 2012; Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011). Everyday experience suggests that memory must at some point become a significant factor in visual search, so the effectiveness of memory may depend on the specific conditions of the experiment. In this respect, there are clear differences between paradigms that directly test whether incidental encoding during visual search or scene viewing leads to formation of memory and the quality of these representations (e.g., Draschkow, Wolfe, & Võ, 2014; Tatler & Tatler, 2013; Williams, Henderson, & Zacks, 2005) and those that studied memory indirectly through facilitation of visual search by memory representations (e.g., Castelhano & Henderson, 2007; Hollingworth, 2012; Võ & Wolfe, 2012). In the current study, our focus is on the latter question: how memory influences gaze allocation in a scene.

There are many differences between experiments with 2D images and ordinary experience in natural, immersive three-dimensional (3D) environments, even when those images are taken from realistic scenes (Chrastil & Warren, 2012; Hayhoe & Rothkopf, 2011). Conventional paradigms often entail very brief exposures to a large number of images that are usually scaled to fit the display. As a consequence, the nature of such exposure substantially differs from daily visual experience, where we are immersed in a relatively small number of environments for longer durations. Spatial learning in 3D environments is also more active in several aspects: For example, movement is self-initiated and accompanied by proprioceptive and vestibular feedback; subjects make active decisions and allocate attention based on the constraints of the task and the structure of the environment (Chrastil & Warren, 2012). These components of active behavior are rarely possible in experiments that attempt to understand scene learning and visual search using two-dimensional (2D) stimuli. In addition, whole-body motion in 3D environments enables the parallel development of both dynamic egocentric (observer-centered relationships between objects and human observer) and allocentric (world-centered representations of object-object relations) representations of the environment. In 2D settings, however, dynamic egocentric representations are not possible (Burgess, 2006; Farrell & Robertson, 1998; Mou, McNamara, Valiquette, & Rump, 2004; Waller & Hodgson, 2006).

Task structure in the real world is also rarely similar to that captured in traditional 2D static paradigms. In this respect, accumulating evidence has shown the strong impact of task goals on attentional deployment. In the context of natural behavior, fixations are directed almost exclusively to regions relevant to behavioral goals (Castelhano, Mack, & Henderson, 2009; Hayhoe, Shrivastava, Mruczek, & Pelz, 2003; Jovancevic-Misic et al., 2006; Land, 2004; Rothkopf, Ballard, & Hayhoe, 2007). The intimate connection between task demands and gaze implies that task may determine the specific information that is attended and encoded in memory. However, there is still ongoing debate on this issue. Võ and Wolfe (2012) demonstrated performance improvement through repeated search for the same sets of targets, suggesting that spatial information encoded during task-relevant experience leads to a benefit relative to semantic guidance. On the other hand, a number of findings suggest that fixated objects are encoded in memory even when they are irrelevant to the current task (Castelhano & Henderson, 2005; Hollingworth, 2012; Williams et al., 2005). These two views may not be inconsistent, as task may prioritize the selection of information to be incorporated into memory, whereas task-irrelevant objects may also be encoded, although with reduced probability (Tatler & Tatler, 2013). Still, how task and memory interact to modulate deployment of attention during ongoing behavior remains unresolved.

There have been some recent attempts to investigate visual search in more ecologically valid conditions. Kit et al. (2014) showed that visual search performance in an immersive virtual environment improves rapidly over repeated search episodes and that memory for object locations was maintained over 3 days. Using real-world environments, Mack and Eckstein (2011) had participants search for objects on the tabletop and demonstrated that co-occurrence of objects, which is part of our priors resulting from daily experience, can serve as a contextual cue for guiding search. Also, with a similar tabletop search task, Howard, Pharaon, Körner, Smith, and Gilchrist (2011) found that objects incidentally fixated on a trial prior to search were found more efficiently when they became targets. Tatler and Tatler (2013) investigated the effects of task instructions on object memory and attention deployment in real-world settings. They found that instructions to

memorize objects led to better performance compared with free viewing; thus, memory for natural environments is tightly linked to the specific task requirements. In the present experiments, the role of repeated search and incidental learning will be examined further. Jiang, Won, and Swallow (2014) investigated visual search in an outdoor environment and found that whole-body movements influence memory representations by allowing subjects to encode target locations in both egocentric and allocentric frames. Foulsham, Chapman, Nasiopoulos and Kingstone (2014) also suggest that head movement is an important factor in search strategies in real environments. According to these results, it is possible that 2D and 3D environments may indeed lead to different encoding strategies. Therefore, in the present experiment, we attempted a direct comparison of 2D and 3D search contexts.

To capture both the stimulus conditions of natural environments and the task context of natural behavior, our first goal for the current study was to monitor eye movements during search in a naturalistic environment. This allowed us to monitor attention during ongoing behavior and also probe the use of memory during repeated search episodes. In most of the conventional paradigms that have demonstrated memory effects on visual search, a preview image is usually presented and then followed by the search task after a short interval. To simulate this situation more closely, we devised a virtual environment in which subjects initially explored a virtual room and then searched for objects in the different rooms of the virtual apartment while walking around in the environment. Search targets were geometric objects, chosen to decrease the likelihood of semantic association with nearby context and to encourage the use of episodic memory, because guidance from semantic context may reduce the role of episodic memory, as shown in Võ and Wolfe's (2013) experiments. Although previous work indicates that geometric objects would not benefit from a preview as realistic objects (e.g., Castelhano & Henderson, 2007; Hollingworth, 2009; Võ & Henderson, 2010), in our experiment we chose a long preview period that gave ample time for incidental fixations and for encoding scene structure. In the first eight trials, the search targets were different on every trial, allowing evaluation of learning of the context. Following this, each geometric object was a search target on two more occasions, which allowed us to assess how spatial memory developed through repeated search experience. Following the searches for geometric objects, participants were also asked to search for local (nongeometric) contextual objects that had been present as part of the visual context during early trials. Draschkow et al. (2014) found that incidental memory for objects formed during visual search is better than memory from intentional memorization. Thus, evaluating search performance for those contextual objects provides another way to characterize the role of task; that is, do people learn where those objects are located from their experience in the environment, even though they had not been specifically designated as relevant items? The effect on search performance of previous fixations to those contextual objects was examined to determine whether incidental, task-irrelevant fixations contribute to subsequent search guidance.

Because the actual stimulus sequence and task context differ so extensively in traditional 2D paradigms and realistic 3D environments, it is difficult to compare the findings in these two situations, given the quantitative and graded nature of some of the effects. We therefore devised a parallel 2D visual search experiment that was designed to make the stimulus and task conditions as similar as possible to the immersive 3D experiment while maintaining many of the features of previous 2D experiments to better compare the results in a quantitative manner. Our 2D and 3D experiments were similar in many aspects: The task structure and the targets chosen were the same. The 2D search scenes were snapshots taken from the 3D environment. The major difference between both tasks is that head and body movements, and thus active spatial learning, were not part of the 2D experiment. This comparison may therefore provide insights into the importance of active body motion in developing spatial memory.

Methods

3D experiment

Experimental environment

The virtual reality (VR) environment consisted of two rooms, a bedroom and a living room, with a corridor in the middle (Figure 1A), and it was created in FloorPlan 3D V11 (IMSI) and then rendered by Vizard 4 (WorldViz). The dimensions of the virtual apartment were maximized to the space available in our lab while avoiding chances of collisions. Each room was 3 m by 6 m, and the corridor was 1 m by 6 m. Participants wore an nVisor SX111 (NVIS) head-mounted display (HMD) through which they viewed the VR environment, and the HMD is equipped with a ViewPoint EyeTracker (Arrington Research; Figure 1B). The HMD has a resolution of 1,280 × 1,024, a horizontal field of view (FOV) of 102° (in total, 76° each eye), and a vertical FOV of 64°. A HiBall motion-tracking system (thirdTech) was used to track 6 degrees of freedom of head position in the environment at around 600 Hz. The latency for updating the visual display following a head movement was 50 to 75 ms.

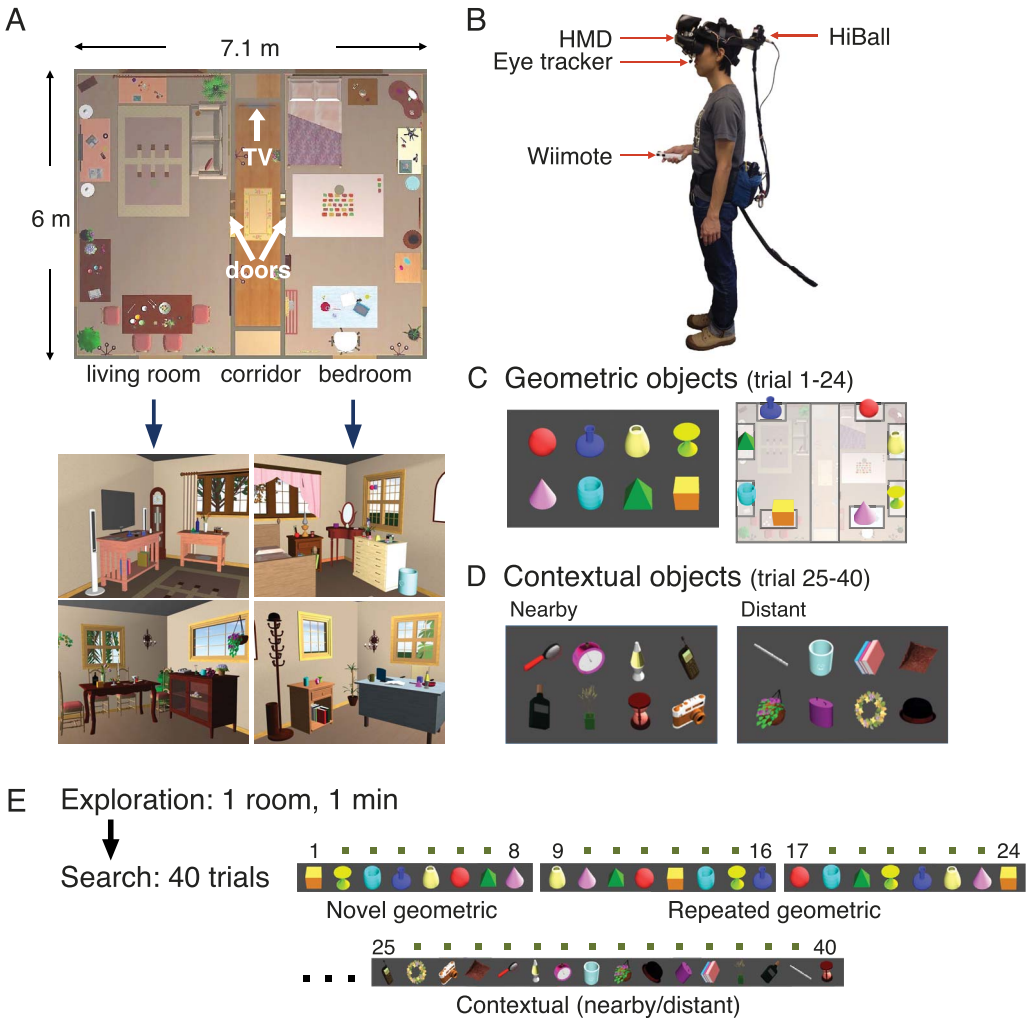

Figure 1. Three-dimensional (3D) experiment. (A) Experimental environment. Top: Birds-eye view of the 3D virtual apartment. The target of each search trial is shown on the TV screen at the end of the corridor between the living room and the bedroom. There are two doors that connect to the rooms. Bottom: Example views of the living room (on the left) and the bedroom (on the right). (B) A participant wearing the head-mounted display equipped with an Arrington eye tracker and HiBall head position tracker, with a second position tracker on the waist (not used here). (C) Left: Eight geometric objects that were search targets from Trials 1 to 24. Right: Schematic of the location of the geometric objects in the apartment. (D) Eight nearby and 8 distant contextual objects that were search targets from Trials 25 to 40. Each nearby contextual object is on the same surface as a geometric search target, and the distant ones are on surfaces different from the geometric search targets. (E) Tasks: Exploration and search. An example trial sequence was shown. Trials 1–8 were novel search trials for geometric objects. The same sets of objects were repeatedly searched for in two additional blocks from Trials 9–16 and 17–24. From Trials 25–40, search targets were a mixture of nearby and distant contextual objects. Trial sequences were randomized across subjects.

The position of the left eye was tracked by the eye tracker at a sampling rate of 60 Hz and an accuracy of about 1°. The eye tracker was calibrated prior to the beginning of the experiments. Because the helmet can shift as the subjects move around the room, the quality of the calibration was checked half way through the experiment as well at the end, using a nine-point (3×3) calibration grid. A poor calibration at the end of the experiment was used as a basis for excluding that subject's data, as were frequent track losses. Recalibration was performed during the experiment if drift was detected. Videos of the eye tracks and the scene display (what participants saw), along with the metadata of the simulation, synchronized for each frame, were stored in an MOV file for later gaze analysis and verification. The metadata include the position of the participants and the objects. Automated analysis was verified using the video records. A Wiimote (Nintendo) was provided for clicking a button when the target was found. Participants were required to be within 1.5 m from the target and looking at the target for the click to end the trial. This was done to

prevent participants from pressing the button without actually finding the target. Once the Wiimote click was detected, a "Trial Done" message appeared on the screen, and participants could proceed to the next trial.

Targets

Target objects subtended approximately 2° to 2.5° of visual angle on average when they were viewed from the entry of the room to the center of room (targets were usually found by the time participants reached the center of the room). There were two types of target objects in the search task: geometric objects (Figure 1C) and contextual objects (Figure 1D). The early search targets were eight geometric objects with homogenous colors. Each of them was placed on one of eight different pieces of furniture (e.g., desk, dining table, side tables, dresser, TV table), four in each room (Figure 1C, right). Later search targets were realistic contextual objects that were continuously present in the apartment and were part of the context during searches for geometric objects. Eight of the contextual targets were nearby (i.e., on the same surface as) the geometric targets that were previously searched for, and the other eight were distant (not on the same surface and could be anywhere in the room). The set of eight geometric objects was searched for in three successive trial blocks, which comprised Trials 1 to 24. The set of 16 contextual objects was searched for only once (Trials 25 to 40).

Procedure

The experiment started with participants moving from the corridor to one room and exploring freely for 1 min to familiarize themselves with the room (see Figure 1E). The explored room was counterbalanced across subjects. The unexplored room served as within-subject control for the effect of pre-exposure in the analysis. Participants were also randomly assigned to either explore the room while geometric targets are absent (context pre-exposure group) or explore while they are present (context-plus-targets pre-exposure group). Thus, half of the participants were only pre-exposed to the context of the room prior to search. Note that the context here includes everything surrounding the geometric objects (including nearby or distant contextual items). The other half of the participants were pre-exposed to both context and targets.

After exploration, participants conducted 40 search trials. At the beginning of each trial, participants returned to the corridor from whichever room they were in and approached the TV screen on the wall at the end of the corridor, which showed an image of the search target for that trial. The participants then had to decide which room to enter to locate the target, and they were allowed to freely traverse between both rooms until the target was found. The rooms had doors that automatically opened when participants were close to the entrance; therefore, participants could not see most of the objects in the room until they had actually entered it. This was done to simplify the analysis of search performance. The trial order was randomized within each block and across participants (an example trial sequence is shown in Figure 1E). The targets of two consecutive trials were never the same. Following three repetitions of the geometric target search trials (total 24 trials), each of the 16 contextual targets was searched for once. The order that nearby and distant contextual targets were searched for was mixed and also randomized across participants.

Gaze analysis

A data file that contained the positions of head and eyes and all objects was generated from reconstruction of the experimental environment in Vizard. The eye position data were analyzed with an automated program developed in house. The data were first preprocessed by a median filter to remove the outliers and then an averaging filter to smooth the signals. A moving window of three frames was used for both types of filtering. The next step was to segment the data into fixations and saccades. The algorithm identified a fixation when the eye movement velocity fell below 60°/s for a period of at least 100 ms. Note that this relatively high velocity threshold is used because of the presence of low-velocity vestibular-ocular reflex movements that add to eye velocity during head movements. Consecutive potential fixations were combined if they were less than 1.5° apart in space and less than 80 ms separated in time. Brief track losses during a fixation were ignored if the object being fixated was identical before and after track loss. In the reconstruction, the objects being fixated were then determined by the program using an 80×80 pixel window (approximately $5° \times 5°$ visual angle) centered at the point of gaze on each frame. This allowed the projection of gaze location on the 2D space of each frame of the video. The window used is relatively large partly because of possible calibration errors and drift during the experiment but also because the target was easily visible even when the participant was not directly fixating it. This is relevant because in natural vision, humans often adopt a "good enough" strategy and do not make corrective saccades unless there is a large error. The eye data were then segmented into trials. The starting point of the trial was defined as first room entry from the door of either room of choice. The end of the trial was defined as the time when the participants made the first fixation to the target of that trial, without making further fixations to other locations until they pressed the

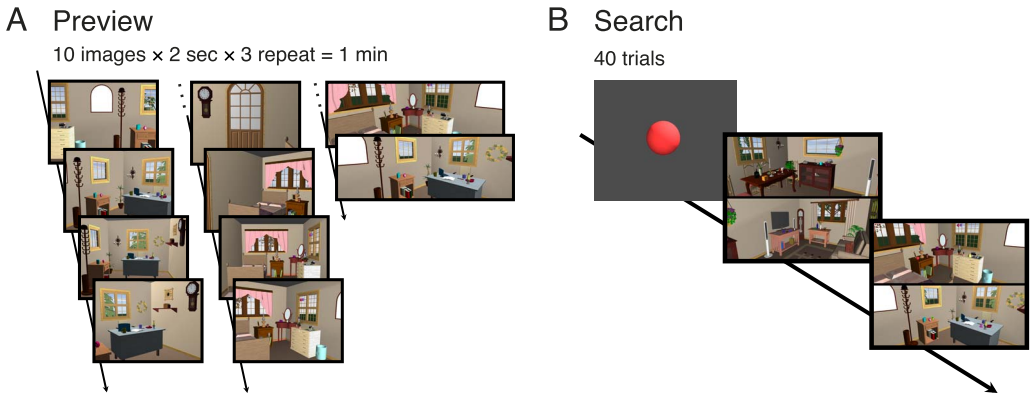

Figure 2. Two-dimensional experiment. Participants viewed images on a computer monitor and performed visual search. (A) Preview images of bedroom. Preview images were repeatedly presented three times, 2 s each image, for a total of 1 min. Note that there was a fixation screen (not shown here) shown in between each preview image in the actual experiment. (B) The search trial started with a target object image displayed on the screen, then participants pressed left or right keys to see the image of the living room or bedroom until the target was found. Each room image is composed of the left panoramic view of the room on the top and the right panoramic view on the bottom.

Wiimote button. If other fixations intervened, as sometimes happens when subjects are unaware they have fixated on the target and continue the search, the next fixation to the target that satisfied the criterion was chosen. This was done to avoid the timing variability produced by the fact that participants had to approach the target before pressing the button (which they often did while maintaining fixation on the target). First fixation on the target then reflects the end of visual search more precisely than the Wiimote press. To analyze the fixations to surfaces containing geometric targets, boxes that were 40 cm above and 20 cm below all those surfaces were added to the environment during the reconstruction. The number of fixations that fell into those boxes was calculated separately from fixations to objects to determine the percentage of fixations directing to the relevant surfaces.

Participants

Forty-two students from the University of Texas at Austin participated in the 3D experiment. The experiments were approved by the University of Texas Institutional Review Board (IRB: 2006-06-0085). All participants had either normal or corrected-to-normal vision, provided written informed consent, and received experimental credit or monetary compensation upon completion. Six participants skipped one or more trials or did not complete the entire experiment; therefore, their data were excluded. Of 36 participants who were included, 18 did not have reliable eye-tracking data, with frequent track losses and drift (as is commonly found in this particular HMD/eye-tracking system, given the heaviness of the helmet and cable and the awkward geometry of the eye tracker). Because this made it hard to identify the time that the target was located, their data were also excluded for the calculation of search time and search fixations. Data from the 18 participants who had good eye tracks were included in the analysis of search time and fixation counts in the 3D experiment. For the analysis of probability of making correct room choice in 3D, data from all participants were included.

Data analysis

All analyses of recorded data were performed using custom-written programs in Matlab. The eye position data were used to calculate two measures of search performance: search time (time spent to locate targets) and number of search fixations to locate the targets. For each trial, the search time and search fixations in both the correct room and the incorrect room were calculated. Data that were three standard deviations away from the mean of each trial were excluded from the analysis. Statistical methods are described in the Results section. For all statistical tests, an alpha level of 0.05 was used. Post hoc analyses were conducted by bootstrapping.

2D experiment

Stimuli and apparatus

To parallel the exploration experience in the 3D experiment, a series of snapshots taken from the 3D apartment rendered by Vizard in desktop mode were used for the preview phase, to parallel the pre-exposure to one of the rooms in the 3D experiment (Figure 2A). There were 10 preview images: eight overlapping views taken by moving the camera to locations that produced views that covered the entire

room and were similar to those seen by subjects moving in the 3D environment, plus two panoramic views (one for the left side of the room and the other for the right side of the room) for each room. Two versions of those images were generated for preview: geometric-object-absent and geometric-object-present, to parallel the exploration in 3D. The angles from which snapshots were taken were fixed between versions. The search images were of both rooms, although only one of them was visible at a given time. The images were composed of two parts: a left panoramic view of the room that appeared on the top part of the screen and a right panoramic view of that same room that appeared on the bottom (see rightmost part of Figure 2A). This arrangement was chosen to mimic the 3D experiment, in which subjects can only see one room at a time and also to add the need for a gaze shift to inspect a different part of a given room. The images were presented on a 24-inch LCD monitor (resolution 1,920 × 1,200) refreshing at 60 Hz that was placed 70 cm away from the participants. Scene images spanned 33.5° (width) × 23.3° (height) of visual angle on average. Target objects subtended approximately 1.4° of visual angle on average (about 1° for geometric objects). Experiments were conducted using a program written in Matlab, with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and the Eyelink toolbox (Cornelissen, Peters, & Palmer, 2002) extensions. The position of the right eye was monitored by an Eyelink II eye tracker (SR Research) sampling at 250 Hz. Participants placed their heads on a chin rest and were asked to hold their head position constant. The eye tracker was calibrated with a nine-point grid before each session, and drift correction was performed before each trial. A keyboard was provided to participants to choose which of the two rooms to view during the search trials by pressing the left or right arrow keys.

Procedure

Before the main experiment started, a practice session of eight trials was given to ensure that participants knew how to use the keyboard to switch room images. During this practice session, participants searched for eight rendered objects that were not used later in the main experiment and that were placed in two outdoor scenes in the virtual environment. For each trial, the event sequence was the same as for the search task in the main experiment: The target was shown on a black screen, and participants pressed buttons to switch among images for search and hit the space bar when the target was found (detailed below). For the main experiment, the task structure was the same as in the 3D experiment: a preview phase followed by a search phase. First, participants previewed images of one room for 1 min, which paralleled the exploration session in the 3D experiment (Figure 2A). To simulate the viewing experience in the 3D experiment, 10 preview images (successive ones were overlapping with each other) were sequentially presented to the participants for three repetitions, 2 s each image, for a total of 1 min. A black screen with a fixation cross in the center was shown for 0.5 s in between each preview images. Then 40 search trials followed (Figure 2B). The search targets and trial structure were the same as those in the 3D experiment. A target object was presented at the center of the computer screen, and participants pressed the left or right arrow key on the keyboard to see either the image of the living room or the bedroom. Then they hit the space bar when the target was found, and the trial ended if the participants were viewing the correct room. Eye-tracking data were recorded simultaneously. The last object fixated was used to check if participants actually found the target later in the analysis, and the data of the trial were excluded if the last fixated object was not the target. The trial was terminated if the target was not found in 30 s after onset of the first room image, and the data of that trial were excluded.

Participants

Twenty students from University of Texas at Austin participated in the 2D experiment. The experiments were approved by the University of Texas Institutional Review Board (IRB: 2006-06-0085). All participants had either normal or corrected-to-normal vision, provided written informed consent, and received experimental credit or monetary compensation upon completion.

Data analysis

The EyeLink Data Viewer 2.11 (SR Research) was used to define regions of interest (approximately 5° in diameter) on each image and to obtain fixation counts and fixation durations on each image and within each trial. The data were then analyzed in Matlab as in the 3D experiment. Similar to the data analysis of the 3D experiments, the starting point of a trial was when the first room image showed up on the screen; the end of the trial was when participants made the first fixation to the target, without further fixating other objects until they pressed the space key to end the trial. Data that were three standard deviations away from the mean of each trial were excluded from the analysis. Fixations to relevant surfaces were also analyzed by counting fixations that fell within 2° of the surfaces.

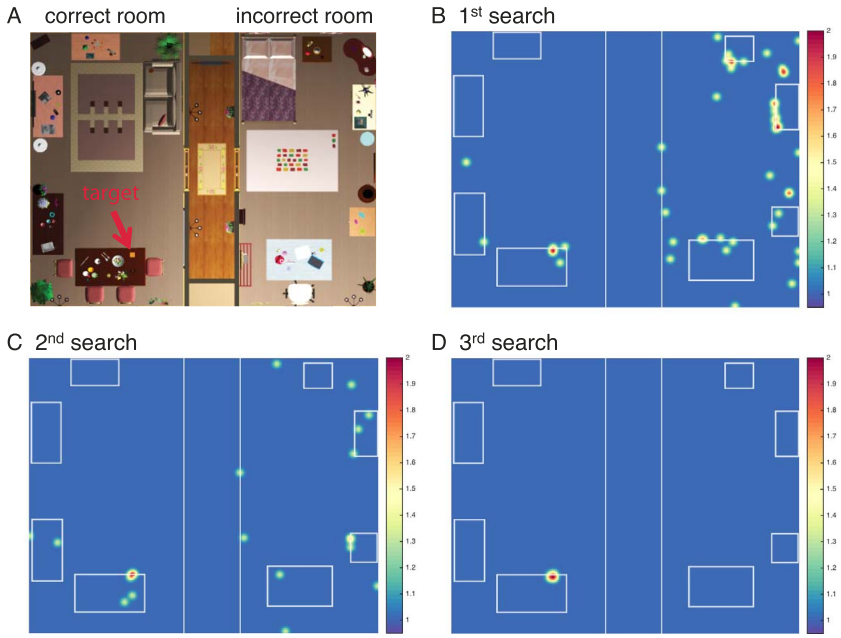

Figure 3. Spatial distributions of fixations in the 3D environment. (A) Birds-eye view of the environment. The red arrow indicates the location of the target, an orange cube. (B–D) Heat maps of fixation counts across space are generated from the fixation data of a subject during the three searches for the orange cube (Trial 2, Trial 10, and Trial 20). Note that the participant searched for other geometric objects in the intervening trials. The walls that separate the rooms from the corridor, as well as the outline of the surfaces that contain geometric targets, are marked with gray lines. For these three trials, the correct room, the room that contained the target, was on the left (living room), and the incorrect room was on the right (bedroom).

Results

To quantify visual search performance, we calculated search time and number of search fixations from the time of first room entry until the target was located. To take accuracy of room choice into account, search time and search fixations were calculated separately for the room containing the target on that trial (the "correct" room) as well as for the incorrect room. Percentage of correct room choice, fixations to relevant surfaces, and incidental fixations were also analyzed.

Effects of repeated search

To illustrate typical performance, spatial distributions of fixations of a subject over three separate searches for the same target are presented in Figure 3. Reduction of search fixations has been observed in both correct and incorrect room over experience. There is also a tendency to restrict fixations to surfaces that contain geometric objects. Performance of all subjects in the first 24 trials with geometric objects is shown in Figure 4. Number of fixations and total search time show similar trends (Figure 4A, B). Overall, performance improved quickly, with the biggest improvement being the reduction of time and fixations spent searching in the incorrect room, especially in the 3D experiment, as shown in Figure 4B. Search improved in the first eight trials, even though the targets were different objects on each trial, suggesting that some aspects of learning the environment produced more efficient search. This could be attributed to either incidental fixations on objects that were targets in later trials, or greater familiarity with the global room structure, or from restricting regions searched to the surfaces containing the targets. In Figure 4C, the data in Part A and B of the figure are grouped into the first, second, and third episodes of eight-trial blocks. Repeated searches for the same target led to reduced fixations and search time in both 2D and 3D in the correct room, where the target was present (see Movie 1 for examples from the 3D experiment). This suggests that subjects rapidly learned the location of the targets in the room once the targets had been searched for. Figure 4C also shows the magnitude of the improvement in the incorrect room (target absent) in the first episode, where the number of fixations dropped from 11 to 3 by the second episode. In a separate analysis, omitting trials in which subjects did not enter the incorrect room did not affect this trend (plot not shown here); that is, there is a rapid drop of fixation count and search time in the incorrect room. Another notable feature of Figure 4C is that although the search time in 2D and 3D was comparable in the correct room

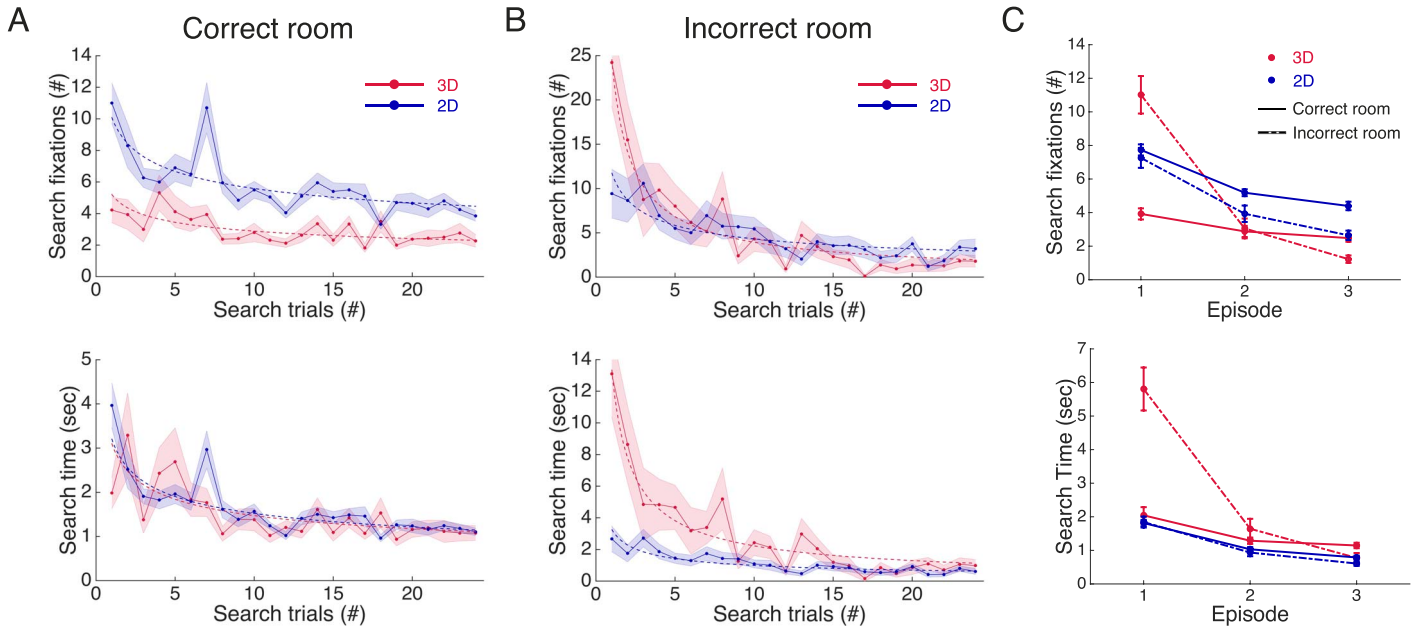

Figure 4. Search performance for geometric objects across search episodes and search trials for 3D (red) and 2D (blue) experiments. (A) Number of search fixations (top) and search time (bottom) across 24 trials once in the correct room. Data from two groups of participants (context pre-exposure group and context-plus-target pre-exposure group) were collapsed within the 2D experiment and 3D experiment. Solid lines show the data series and dotted lines show the exponential model fit. (B) Number of search fixations (top) and search time (bottom) in the incorrect room. Note the different scales for correct room and incorrect room in (A) and (B). (C) Number of search fixations (top) and search time (bottom) in the correct room and the incorrect room across three search episodes. Data represent mean ± *SEM*.

(around 1–2 s), there were roughly twice as many fixations in 2D as in 3D (around six vs. three fixations).

A three-way mixed analysis of variance (ANOVA) was conducted to examine the effect of 2D versus 3D, correct versus incorrect room, and search episode on the number of search fixations and search time. The details of the statistical analysis are shown in Table 1. The improvement over episodes in Figure 4C is revealed by the significant effect of episode. This is reliable in both 2D and 3D. The significant three-way interaction reflects the big drop in the first to second episodes in 3D in the incorrect room. To assess whether baseline performance and rate of improvement are different between 3D and 2D experiments, we used nonlinear mixed-effect modeling (NLME; Pinheiro & Bates, 2000) to generate parameters that fit an exponential function to our search fixation and search time data across the trials (see Brooks, Rasmussen, & Hollingworth, 2010, for an example of using NLME to analyze data from contextual cueing experiments). Compared with traditional ANOVA, adopting NLME has the following advantages in our study. First, NLME treats time (in our case, search trials) as a continuous factor, rather than categorical as in ANOVA, which increases statistical power. Second, NLME allows a better characterization of the learning function, as an exponential function. Third, within-subject variability is also taken into account in NLME. Once we obtained the parameters, including intercepts and slopes of the exponential function, from NLME, we bootstrapped the difference of distributions of intercepts and slopes to indicate the difference in baseline performance and improvement rate, respectively, in 2D and 3D experiments. The details of this analysis are provided in Appendix A and indicate that in the correct room, fewer fixations were made in 3D than in 2D in the first trial, although neither search time nor improvement rates were different.

Memory representations alter attention allocation

There are several potential reasons for the improvement in performance with increased experience. First, subjects may become better at choosing the correct room. To examine this possibility, the percentage of trials in which the correct room was chosen first, as well as the number of room entries required to locate the target, was calculated (see Figure 5A). There is significant improvement in the correct room choice across search episodes (mixed-model ANOVA), $F(2, 120) = 16.5$, $p < 0.001$ (also see Movie 1 for examples from the 3D experiment), and the probability was greater in 3D than in 2D, $F(1, 60) = 16.08$, $p < 0.001$,

Parameter	Effect	*F*	*p*
Number of search fixations	Experiment	$F(1, 36) = 9.45$	0.004
	Room	$F(1, 36) = 2.4$	0.13
	Episode	$F(1.32, 47.6) = 119.6$*	<0.001
	Room × Experiment	$F(1, 36) = 33.6$	<0.001
	Episode × Experiment	$F(1.32, 47.6) = 4.04$*	0.04
	Room × Episode	$F(1.5, 54.8) = 39.86$*	<0.001
	Room × Episode × Experiment	$F(1.5, 54.8) = 22.69$*	<0.001
Search time	Experiment	$F(1, 36) = 37$	<0.001
	Room	$F(1, 36) = 19.76$	0.001
	Episode	$F(1.16, 41.8) = 100$	<0.001
	Room × Experiment	$F(1, 36) = 25.81$	<0.001
	Episode × Experiment	$F(1.16, 41.8) = 21.26$	<0.001
	Room × Episode	$F(1.3, 46.7) = 33.02$	<0.001
	Room × Episode × Experiment	$F(1.3, 46.7) = 27.42$	<0.001

Table 1. Effects of experiments (3D vs. 2D), room (correct vs. incorrect), and search episode (1–3) on search performance for geometric objects: three-way mixed ANOVA for data in Figure 4C. *Mauchly's test of sphericity indicated violation of assumption of sphericity for the effect, and therefore, a Greenhouse-Geisser correction was applied to correct the degrees of freedom, which affects the *p* values.

and no interaction was found, $F(2, 120) = 1.59$, $p = 0.21$. There is also a reduction in number of room entries across episodes (mixed-model ANOVA), $F(2, 72) = 12.19$, $p < 0.001$. Overall, participants make slightly more (about 0.2) room entries in 2D than 3D, $F(1, 36) = 15$, $p < 0.001$. These findings also indicate that subjects tend to use memory more in 3D than in 2D, perhaps as a result of the higher energetic or time cost of changing rooms.

Another possibility is that they learn the relevant parts of the room to look at. We analyzed the proportion of fixations that were restricted to potential target locations (i.e., the surfaces in which geometric objects were placed (Figure 5B). Search fixations directed to the relevant surfaces increased (see Movie 1 for examples from the 3D experiment) from about 60% to 87% in 3D (from the first to the last trial) and from about 66% to 78% in 2D. The improvement is significant in both experiments (one-way ANOVA), 3D: $F(23, 420) = 1.83$, $p = 0.01$; 2D: $F(23, 479) = 1.86$, $p = 0.009$. The difference between 2D and 3D was not significant, $t(46) = 1.24$, $p = 0.22$. Note that even on the first trial, subjects are biased to restrict search to surfaces like tables. Thus, search benefits from experience by excluding irrelevant parts of the scene, including the incorrect room and irrelevant regions in the rooms, during the search process.

Pre-exposure effects

To assess the effects of pre-exposure on search performance on the first trial, we first looked at the effect of pre-exposure on search time and search fixations. This is presented for the correct room (target

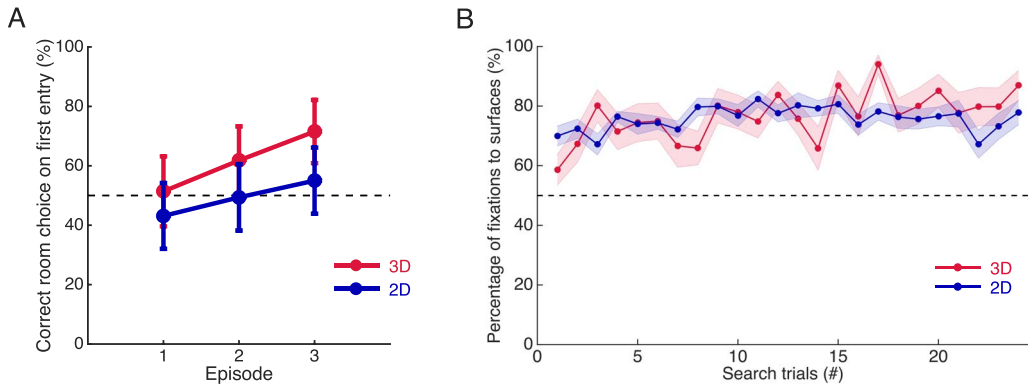

Figure 5. Performance improved as a result of excluding irrelevant parts of the environment. (A) Percentage of correct room choice on first room entry across search episodes in the 2D experiment and 3D experiment. (B) Percentage of fixations made to surfaces that contained geometric targets in the 2D and 3D experiment. Data represent mean ± *SD* in (A), mean ± *SEM* in (B).

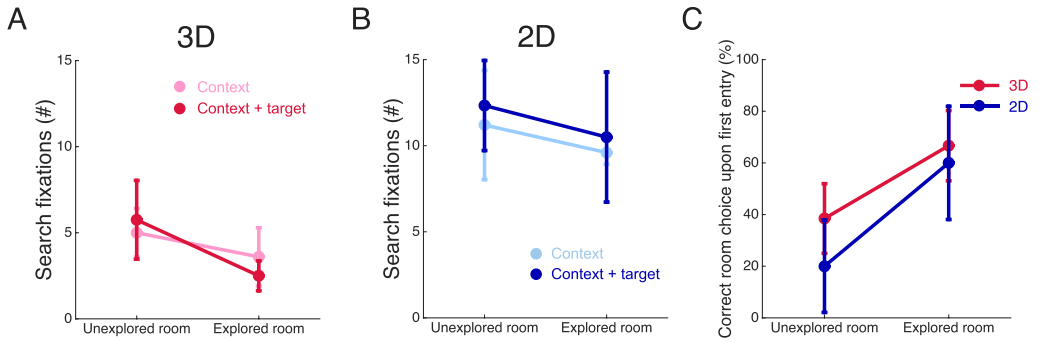

Figure 6. Effects of pre-exposure on search performance. (A) Search fixations when the target of the first trial was in the unexplored room and the explored room, in the group that was pre-exposed to room context only, and the group that was pre-exposed to both the context and the targets in a room in the 3D experiment. (B) Same as (A) but in 2D. (C) Percentage of correct room choice upon first room entry on the first trial when the target was in the unexplored and the explored room, in the group that was pre-exposed to both the context and the targets, in 2D and 3D. Data represent mean ± *SEM* in (A) and (B), mean ± *SD* in (C).

present) in Figure 6A, B. Pre-exposure was either to the room only or to the room plus targets. Although it appears that search fixations were less numerous in the room that participants explored, at least when exploration included the targets, there were no significant differences in either 2D or 3D: 2D, room: $F(1, 19) = 0.42$, $p = 0.53$, targets: $F(1, 19) = 0.15$, $p = 0.7$); 3D, room: $F(1, 17) = 1.93$, $p = 0.19$, targets: $F(1, 17) = 0.01$, $p = 0.91$. However, because we are looking only at the first trial, the statistical power of the comparisons is very poor. Therefore, these results are inconclusive. However, there may be some effects of pre-exposure to both room and targets on the probability of choosing the correct room for the initial search. These data are presented in Figure 6C. Seeing targets during pre-exposure seems to facilitate selection of the correct room, although the difference between rooms does not reach significance as a result of the small sample size (Fisher's exact test, 3D: $p = 0.24$; 2D: $p = 1$).

In summary, the primary conclusions from these data are (a) with only three repeated searches of the same objects, performance improved substantially, indicating rapid encoding in spatial memory; (b) performance in 2D and 3D is comparable in this respect, although number of fixations in the correct room was overall lower in 3D (discussed below); (c) search time in the incorrect room in 3D is very large for the first few trials, perhaps reflecting a reluctance to change rooms because of the energetic cost; (d) a significant component of the improvement with experience is the reduced time searching in irrelevant locations, such as the incorrect room or the regions less likely to have targets; and (e) there is some indication of an advantage from pre-exposure to the scene, but the weak power of the statistical tests does not allow any firm conclusions.

The role of task relevance: are local contextual objects learned?

We demonstrated that when an object is the target of search, subjects learn which room it is in, what parts of the room might be relevant, and the location of the target in the room. We then examined whether subjects learn the location of other objects in the environment that have not been explicitly searched for. While searching for the designated geometric target, participants made many fixations in the room, and the exposure to the search environment was substantially longer (>10 min) than in the pre-exposure period (1 min). In the second phase of the experiment (Trials 25–40), objects that were part of the context during early searches were chosen as targets. It is possible that nearby contextual objects were part of the memory representations formed for targets in the early searches. Here we examined whether the locations of these contextual objects were learned from early search experience. We also compared search for nearby contextual objects with more distant objects. Figure 7 shows search fixations for nearby and distant contextual objects compared with that for geometric objects, both in the first and the later episodes. Search time and fixation count data show the same trends, so only search fixation count data are shown here. The significant difference between novel and repeated search for geometric objects has been shown in the results described above (Figure 4), so our interest now lies in the comparison of the search performance for contextual objects versus geometric objects.

These data are shown in Figure 7. In both 2D and 3D, search for contextual objects was no more efficient than a novel search for geometric objects. Bootstrapped distributions were used to determine the significance of pairwise differences in the measures of search performance, as the samples are not independent and we thus

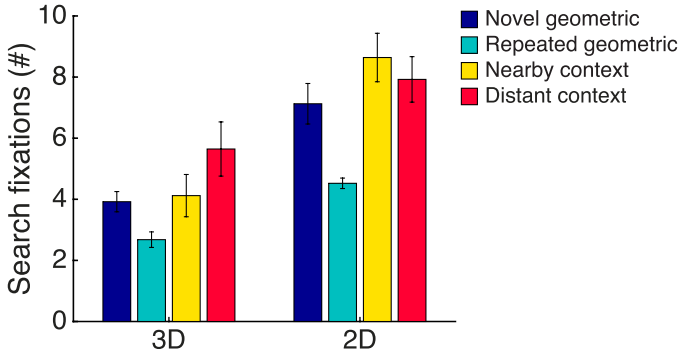

Figure 7. Number of search fixations for novel (Trials 1–8) and repeated (Trials 9–24) search for geometric objects and for nearby and distant contextual objects in the correct room in 3D experiment and 2D experiment. Data represent mean ± *SEM*.

did not use ANOVA first. For detailed description of the statistical analysis, see Appendix B. Our data indicate that participants learned to search for specific items in the room, and the spatial learning that occurred during this process may not be sufficient to support searching for a different set of items, possibly because they were not relevant for the task early on. However, subjects may still learn more general aspects of the room structure even if they do not learn the locations of specific objects.

Contribution of incidental memory to search

It is unclear whether the failure of contextual objects to benefit from experience is a consequence of not being fixated or being fixated but not remembered. We therefore investigated the effect of incidental fixations, which are fixations made to an object before it becomes a target. This also allowed us to examine the extent to which search performance can be attributed to memory built up during task-irrelevant incidental fixations. That is, do fixations on irrelevant objects, which are nearby previously searched items, help later searches for those objects? Incidental fixations were calculated for each nearby and distant contextual object including all trials that occurred before the first time they were searched for. For the nearby contextual objects, 84.4% of them were incidentally fixed at least once before searched for in 3D, 90% in 2D; for the distant contextual objects, 53.5% of them were fixated before becoming search targets in 3D and 77% in 2D. Thus, in general, contextual objects are fixated more in 2D, perhaps because of the smaller visual angle of the display. We plotted histograms separately for nearby (see Figure 8) and distant contextual objects and found that there was no discernable effect of incidental fixations for the distant contextual objects, consistent with the reduced frequency of incidental fixations on those objects.

The frequency distributions of search fixations for nearby contextual objects, given the different number of incidental fixations to those same objects during prior searches, were plotted for both 3D and 2D (Figure 8A, B, top). Cumulative distributions are also shown here (Figure 8A, B, bottom). Data for one or more incidental fixations are combined in Figure 8A and 8B because one to three incidental fixations were most common, and the corresponding histograms of search fixations do not differ much. The frequency distribution of number of search fixations in the case in which there had been no prior fixation on the target is compared with the case in which one or more incidental fixations had been made on the target. Two main findings emerge: First, there is some indication that incidental fixations shift the distribution of search fixations leftward slightly, especially in 2D. However, there is considerable variability. Despite prior incidental fixations to some objects, many targets still required multiple search fixations. Most of the time, the nearby contextual objects were fixated during early searches, yet fixation does not seem to guarantee faster search (or memory for the locations of objects). An alternative way of plotting this data is shown in Figure 8C and 8D, which present search fixations for a given number of incidental fixations. These plots show no clear trend of the effect of incidental fixations (a measure of goodness-of-fit of the raw data, R^2, is 0.001 for both 3D and 2D data). Thus, it appears that the cumulative distributions in Figure 8A and 8B are more suggestive in revealing a small effect of incidental fixations.

Discussion

The goal of this study was to examine how memory affects gaze allocation in natural environments. We investigated how search changed as a function of experience, what components of the scene guided search, and whether performance was similar in 2D and 3D versions of the experiment that were parallel in task structure. Experience in the search task led to rapid improvement in both 3D and 2D. In the first eight trials, search improved substantially, even though targets were different on each trial. A large part of this improvement was reduction in the time spent in the incorrect room, especially in 3D, where it dropped dramatically in the first three trials. Learning in the first eight trials was also accompanied by increased probability of choosing the correct room and fewer fixations to irrelevant regions within the rooms. Thus, some global aspects of the scene context aided search in the first eight trials, despite our failure to find evidence for

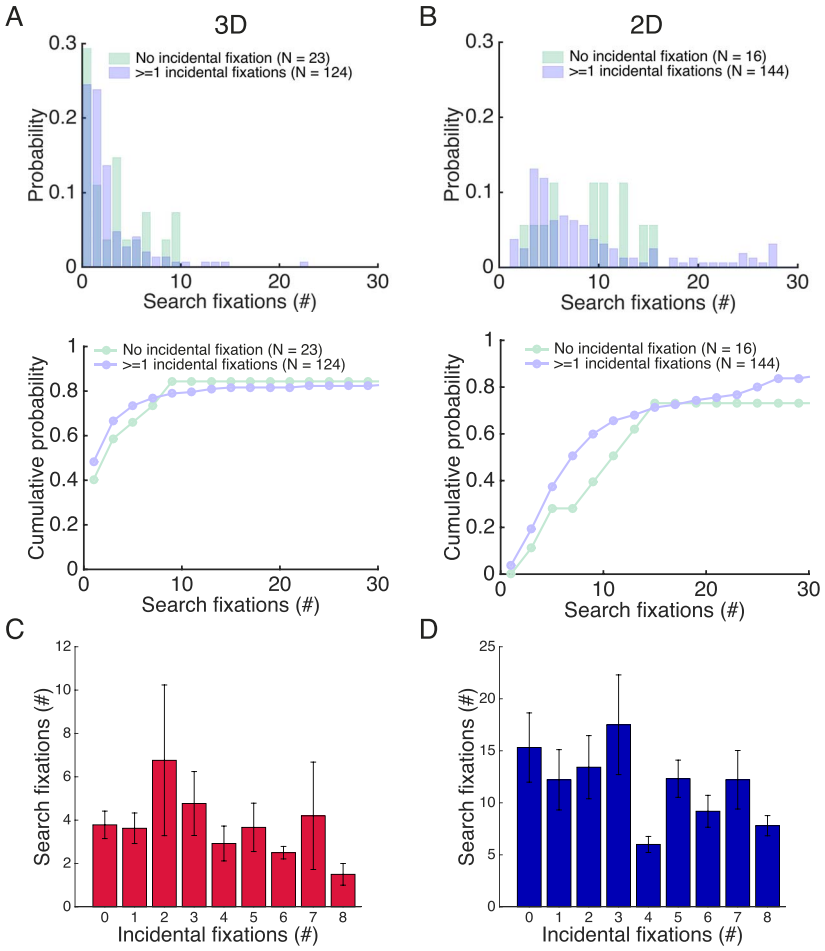

Figure 8. Incidental fixations do not always facilitate search. Distributions of search fixations in the correct room when no incidental fixations were made or when at least one incidental fixation was made prior to search trials in (A) 3D experiment and (B) 2D experiment. Graphs on top show the original distributions and graphs on the bottom show the cumulative probability distributions of search fixations. In (A) and (B), probabilities or cumulative probabilities for making more than 30 search fixations were not shown so that the difference between the distributions could be seen easily. (C) Number of search fixations as a function of number of incidental fixations made in 3D experiment. (D) Same as (C) but for 2D experiment.

an advantage of the 1-min pre-exposure to the context. When search targets were repeated in the second and third episodes, subjects improved over only three repetitions, indicating that locations are stored in spatial memory. Despite extensive experience in the environment during search for geometric objects, performance of search for contextual objects was no better than early search of geometric objects. Extending this finding, we observed very little effect of incidental fixations on subsequent search trials. Thus, learning the location of specific objects appears to depend primarily on previous search for that object. Finally, 2D and 3D search were similar in most respects, with the primary differences being the better selection of the correct room to search in 3D as well as more initial fixations in the incorrect room, indicating higher utility of memory in 3D, which is likely a consequence of higher energetic cost of whole-body movements.

Memory for context

Repeated search

There are a number of factors that might have led to the improvement in the first eight trials. One factor is that subjects increasingly restricted fixations to the four surfaces where the geometric targets were located. This kind of advantage, driven by memory for the scene structure, was suggested by Wolfe et al. (2011), who found that in realistic scene images, the number of items fixated is restricted by prior knowledge of the scene. This was proposed by Neider and Zelinsky (2008), referred to as *functional set size*. Interestingly, in our experiments, this strategy was present even in the first trial, suggesting that prior knowledge of typical room structure guided search. Another factor leading to improvement in the first eight trials was the reduction of time and fixations in the incorrect room, where the target was absent, most notably in 3D in the

first few trials (Figure 4B). Once a subject had chosen the incorrect room in 3D, the cost of leaving and searching the other room is likely to be higher, both in energy and in time, compared with 2D, and this may have led to the longer initial search times. With experience, subjects were able to reject the incorrect room much more quickly, and by the third to fifth trial, the number of fixations in the incorrect room were very similar for both 3D and 2D. It is not clear exactly what is being learned here. One possibility is increasing familiarity with the global room structure that allows faster search. Another possibility is that subjects were better able to define a generalized search template for colored geometric objects. This latter suggestion is consistent with the result that a 1-min preview that included geometric objects reduced search time in the incorrect room. We also found improvement for both 2D and 3D searches in the first eight trials in the correct room where the target was different on every trial. In Võ and Wolfe (2012, 2013), subjects searched for realistic objects embedded in the scene images. These authors found that search was typically guided by general knowledge of object location within natural scenes when such cues were available; however, when those cues were unavailable, episodic memory of object-scene relationships was used to guide search. This is in line with our finding here. Because we adopted geometric objects as search targets, scene semantics were not as useful for our task, and thus it is likely that episodic memory was used to guide search. The improvement in our case suggests a rapid formation of episodic memories associating targets and the scenes as well as the scene layout through active search experience. However, when contextual objects were searched later on, they did not seem to benefit from scene semantics very much. This might be a result of subjects adopting a search template specifically for geometric objects, which may restrict the extent to which other items are processed even when gaze lands on them. Thus, the nature of the search template may determine what aspects of the context are remembered, by excluding nonmatching objects from memory. This result will be revisited later in the discussion of pre-exposure effects.

Despite the large improvement in the avoidance of the incorrect room, the improvement we found over repeated search episodes in the correct room (where the target is present) is quite small (about one to two fixations) although numerically comparable to the improvement found previously (e.g., Kit et al., 2014; Võ & Wolfe, 2012). One speculation is that the fairly simple layout of objects and potential search surfaces, as well as the relatively small scale of the apartment room (3 m $\times$ 6 m each room), made search in our environment easy enough that visual guidance dominated search once the body and head were directed to the local region that contained the target. A more difficult or energetically costly task would likely recruit memory more (Ballard, Hayhoe, & Pelz, 1995; Solman & Kingstone, 2014; Solman & Smilek, 2012). For instance, a larger space with more items on each search surface or increasing similarity between targets and distractors might increase the effects of memory (Duncan & Humphreys, 1989; Neider & Zelinsky, 2008).

Compared with previous results of contextual cueing with 2D stimuli, the effect of repeated search is bigger and develops faster. With simple array stimuli, Chun and Jiang (1998) showed that the contextual cueing effect became noticeable after five repetitions of search trials. The improvements continue after 15 to 20 repetitions, with a total improvement of 60 to 80 ms. With 2D images of real-world scenes, Brockmole and Henderson (2006b) found that only four repetitions are required to reach maximum benefit, which is 20 times larger than that of simple stimuli. In both of our 2D and 3D experiments, repeated search benefits with only one repetition in the correct room, and the magnitude is about 1 s. Thus, our results also show that the learning of target-scene association is faster in naturalistic environments and scenes.

Pre-exposure effects

We were unable to determine the effect of pre-exposure on search because of insufficient power, although there is a hint from our data that seeing the targets during pre-exposure may promote selection of the correct room on the first trial. We speculate that to see a context pre-exposure effect in complex 3D environments, more extensive and interactive experience may be required to generate robust memory representations. There are several possible reasons that pre-exposure experience in our experiment would not result in much search benefit. First, participants were not informed about the future targets during exploration and thus lacked explicit goals or expectations for the main task. Free exploration, like free viewing, is unlikely completely task free/goal free (see discussion in Tatler, Hayhoe, Land, & Ballard, 2011). Tatler and Tatler (2013) demonstrated that free viewing leads to worse memory recall of objects in real-world environments than when participants were asked to memorize all or a subset of the objects. Even intentional memorization may not lead to a better memory recall of objects than when incidentally encoded during visual search (Draschkow et al., 2014). Second, as mentioned previously, the use of geometric objects minimizes search guidance from the knowledge of the object-scene relations. We chose geometric objects to evaluate the role of episodic memory of object-scene relationships,

because scene semantics determine search performance when search targets are familiar (Võ & Wolfe, 2013). We were interested in examining the effects of longer pre-exposure periods than those typically used (e.g., Hollingworth, 2009; Võ & Henderson, 2010). Because even a relatively long exposure had little effect, a more interactive experience may be required to form episodic memories that are useful for later search. Third, it is also possible that pre-exposure may be useful only for guiding the first few fixations to the potentially relevant parts of the scene (Hillstrom et al., 2012). This is consistent with our finding that a high proportion of fixations were directed to relevant surfaces in the room even on the first search trial,

It is also worth noting that the preview benefits reported in the literature might result from the constraints of conventional paradigms, in which tasks are usually more obvious to the participants even without explicit instructions. This might be caused by the task structure that those paradigms use, within which participants look for a target shortly after the preview scene was presented, and this event sequence occurs for a large number of image-target pairs in a short time. Thus, during each scene preview, there is likely an inherent task for the participants to remember the scene components as much as possible to prepare for a forthcoming search. Varying the nature of the interaction with the environment during pre-exposure may also influence the extent to which memory representations develop and thus generate a more profound effect on subsequent search. One example could be allowing manual exploration of objects in the environment instead of just visually browsing. Providing specific instructions that imply relevance of certain objects may also change the influence of such exposure to later search. For example, Tatler and Tatler (2013) instructed subjects to remember tea-related objects and that led to higher chance of directing fixations to those items and also more fixations made to them compared with the free-viewing condition.

Local contextual objects and incidental fixations

Although memory for some aspects of the context seems to benefit search, search for nongeometric contextual objects in the room was no better than initial search for geometric objects despite extended experience (at least 24 trials) prior to those searches. Early in the experiment, participants might learn that only geometric objects were targets, and thus surrounding objects, even nearby ones that might have received some incidental fixations, were not considered as task relevant until contextual objects became targets. This is consistent with the finding, with simple stimuli, that subjects learn to restrict attention to relevant items with repeated search experience (Kunar, Flusberg, & Wolfe, 2008). Thus, there exists a delicate tradeoff during spatial learning: The more prioritized the relevant information, the less the irrelevant information is processed (Tatler & Tatler, 2013). Together with the findings discussed above that experience leads to increase attention to the room and its relevant parts, the poor performance for local contextual objects suggests that search is primarily guided by memory representations of global components of scenes. In this case, the room and layout of furniture in the room are learned, rather than local components, including the nearby or distant contextual items. This is consistent with Brockmole et al. (2006) in that global context is more critical for search guidance. It is also consistent with the idea of Brooks et al. (2010) that representations of the environments may be hierarchically constructed, although we showed only that different levels of context were differentially affected by experience and did not directly test whether each level is nested within another. Here we extend the finding that the more local level of context is not well represented in our environments. The detailed representations may not be built up unless relevant.

The small effect of incidental fixations on nearby contextual objects supports this account. Fixating an object one or more times does not guarantee more efficient search, as indicated in prior studies (Castelhano & Henderson, 2005; Hout & Goldinger, 2010; Olejarczyk et al., 2014; Williams et al., 2005). Thus, incidental fixations to individual objects may not be the primary contributor to search performance. It is possible that some properties of those objects were remembered through previous experience, but their locations in the environment may not be well represented in memory. This can be attributed to the task effect: Subjects may form a search template that prioritized the geometric objects and deprioritized a detailed representation of irrelevant items—in this case, the local contextual objects, in the environment. Howard et al. (2011) found incidental learning in consecutive trials in real-world search, so the timing of incidental fixations may play an important role in this. The temporal and spatial history of incidental fixations and how they contribute to search need to be further explored to provide more refined insight on this issue.

We also found a greater percentage of fixations to contextual objects in 2D than in 3D. This may be attributed to the overall smaller visual angles between the contextual objects and the geometric objects (by about a factor of two), which may also be related to the more separated distributions of fixations in 2D (Figure 8B). This effect is particularly pronounced for distant contextual objects, because some of them require a head turn to be fixated in 3D (e.g., cushion on the

couch), whereas only small eye movements are required (and allowed) in 2D.

Comparison of 2D and 3D

In general, the results in 2D and 3D are very similar. However, as mentioned above, there were fewer search fixations in 3D than in 2D (Figure 4A). In 3D, a substantial fraction of the search time was taken up by large head movements, during which fixations were not identified, with a consequent reduction in fixation counts. Although this is the primary factor, we cannot rule out differences in saccade detection. When the head is moving, the vestibular-ocular reflex adds to eye velocity, so a higher velocity criterion for detecting saccades was needed. In addition, noise in the 3D signal was counteracted by clumping fixation locations within a 1.5° radius into a single fixation. Thus, smaller saccades are easier to detect in 2D. So although a smaller number of fixations in 3D is most likely a consequence of the need to make large head movements, it is hard to be confident of the precise magnitude of the difference. Eye movements in 2D are less energetically costly, as they are typically smaller and are not accompanied by head movements, and this potentially accounts for the smaller number of fixations in 3D.

The other major difference was the large number of fixations in the incorrect room in 3D for the first few trials. Subjects may have been reluctant to exit the room until they were sure the target was not there, because of the big cost of changing rooms. As discussed above, this might point to one of the important characteristics of experience in the 3D environments: The overhead of moving the body from one room to the other, compared with the ease of looking from one room to another in the 2D experiment, may lead to very different strategies. In 3D, search involves full-body motion whereas only eye movements are allowed in our 2D task. The greater probability of choosing the correct room in 3D is also consistent with the adoption of different strategies.

To further understand the causes of rapid improvement in 3D, we investigated how subjects exclude irrelevant spatial regions at two levels: decreasing entries to the incorrect room and avoiding visual exploration of irrelevant parts of the rooms. At the first room entry after the beginning of each trial, the probability of choosing the correct room to search increased from chance level to about 70% on the third episode in 3D, yet in 2D, this choice remained at chance levels (despite some hints of improvement). In addition, a smaller number of room entries was required to locate the target in 3D than in 2D. This may also reflect the fact that it is relatively easy to switch between rooms by pressing keys in 2D. The rapid increase of probability of directing fixations to potential target locations upon room entry is another fact that accounts for the sharp improvement seen in incorrect room in 3D. Interestingly, even at the first trial, about 60% of the fixations were devoted to relevant locations. This may indicate the tendency to look at surfaces that likely contain objects based on prior knowledge and perhaps also from the presearch exploration experience. It is consistent with the cognitive relevance framework proposed by Henderson, Malcolm, and Schandl (2009) that suggests that attention is directed to regions relevant to the current task goal based on scene knowledge during search in realistic scenes. This is achieved by integrating prior knowledge and task to narrow down regions to be searched for (Kunar et al., 2008; Neider & Zelinsky, 2008; Wolfe et al., 2011) effectively, even with only a glimpse of a scene (Castelhano & Henderson, 2007; Võ & Henderson, 2010). Here we also showed that repeated search experience is an important factor in this process of "cutting down the irrelevant regions from search" in a naturalistic environment.

Conclusions

Our results indicate the importance of task in learning the spatial structure to support visual search. The effect of context is very sensitive to the task-specific nature of prior experience in both 2D and 3D environments. In general, search performance in 2D and 3D environments was quite similar, although body movements in 3D allow stronger guidance from the scene memory and structure. Such guidance is characterized not just by associating target location with global scene structure but also by restricting visits of the eyes and the body to the regions of the scene that are irrelevant to the goal. Thus, memory for global spatial context is important in making search more efficient by directing the body to the relevant scene regions.

Keywords: visual search, memory representations, attention, eye movements

Acknowledgments

We appreciate the support from grants NIH EY05729 and PSI2013-43742. M.P.A. was supported by a UAM-Banco Santander Inter-University Cooperation Project and by a "José Castillejo" International Mobility Grant (Ministerio de Educación, Cultura y

Deporte, Programa Nacional de Movilidad de Recursos Humanos del Plan Nacional de I+D+i 2008-2011).

Commercial relationships: none.
Corresponding author: Chia-Ling Li.
Email: sariel.cl.li@utexas.edu.
Address: SEA 4.128D, Mailcode A8000, Austin, TX 78712.

References

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*, 66–80, doi:10.1162/jocn.1995.7.1.66.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual cueing in naturalistic scenes: Global and local contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 699–706, doi:10.1037/0278-7393.32.4.699.

Brockmole, J. R., & Henderson, J. M. (2006a). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *Quarterly Journal of Experimental Psychology*, *59*, 1177–1187, doi:10.1080/17470210600665996.

Brockmole, J. R., & Henderson, J. M. (2006b). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*, 99–108, doi:10.1080/13506280500165188.

Brooks, D. I., Rasmussen, I. P., & Hollingworth, A. (2010). The nesting of search contexts within natural scenes: evidence from contextual cuing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1406–1418, doi:10.1037/a0019257.

Burgess, N. (2006). Spatial memory: How egocentric and allocentric combine. *Trends in Cognitive Sciences*, *10*, 551–557, doi:10.1016/j.tics.2006.10.005.

Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*, *18*, 890–896, doi:10.3758/s13423-011-0107-8.

Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition*, *12*, 1017–1040, doi:10.1080/13506280444000634.

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 753, doi:10.1037/0096-1523.33.4.753.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*(3):6, 1–15, doi:10.1167/9.3.6. [PubMed] [Article]

Chrastil, E. R., & Warren, W. H. (2012). Active and passive contributions to spatial learning. *Psychonomic Bulletin & Review*, *19*, 1–23, doi:10.3758/s13423-011-0182-x.

Chun, M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *71*, 28–71.

Chun, M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*, 360–365, doi:10.1111/1467-9280.00168.

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*, 613–617.

Draschkow, D., Wolfe, J. M., & Võ, M. L.-H. (2014). Seek and you shall remember: Scene semantics interact with visual search to build better memories. *Journal of Vision*, *14*(8):10, 1–18, doi:10.1167/14.8.10. [PubMed] [Article]

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458, doi:10.1037/0033-295X.96.3.433.

Farrell, M. J., & Robertson, I. H. (1998). Mental rotation and the automatic updating of body-centered spatial relationships. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 227–232, doi:10.1037/0278-7393.24.1.227.

Foulsham, T., Chapman, C., Nasiopoulos, E., & Kingstone, A. (2014). Top-down and bottom-up aspects of active search in a real-world environment. *Canadian Journal of Experimental Psychology*, *68*, 8–19, doi:10.1037/cep0000004.

Hayhoe, M. M., & Rothkopf, C. C. A. (2011). Vision in the natural world. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 158–166, doi:10.1002/wcs.113.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. J. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1):6, 49–63, doi:10.1167/3.1.6. [PubMed] [Article]

Henderson, J. M., Malcolm, G. L., & Schandl, C.

(2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*, *16*, 850–856, doi:10.3758/PBR.16.5.850.

Hillstrom, A. P., Scholey, H., Liversedge, S. P., & Benson, V. (2012). The effect of the first glimpse at a scene on eye movements during search. *Psychonomic Bulletin & Review*, *19*, 204–210, doi:10.3758/s13423-011-0205-7.

Hollingworth, A. (2006). Scene and position specificity in visual memory for objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 58–69, doi:10.1037/0278-7393.32.1.58.

Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition*, *17*, 273–291.

Hollingworth, A. (2012). Task specificity and the influence of memory on visual search: Comment on Võ and Wolfe (2012). *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1596–1603, doi:10.1037/a0030237.

Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception & Psychophysics*, *72*, 1267–1282, doi:10.3758/APP.

Howard, C. J., Pharaon, R. G., Körner, C., Smith, A. D., & Gilchrist, I. D. (2011). Visual search in the real world: Evidence for the formation of distractor representations. *Perception*, *40*, 1143–1153, doi:10.1068/p7088.

Jiang, Y., & Wagner, L. C. (2004). What is learned in spatial contextual cuing—Configuration or individual locations? *Perception & Psychophysics*, *66*, 454–463, doi:10.3758/BF03194893.

Jiang Y. V., Won, B. Y., & Swallow, K. M. (2014). Spatial reference frame of attention in a large outdoor environment. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1346–1357, doi:10.1016/j.biotechadv.2011.08.021.Secreted.

Jovancevic-Misic, J., Sullivan, B., Hayhoe, M. M., Jovancevic, J., Sullivan, B., & Hayhoe, M. M. (2006). Control of attention and gaze in complex environments. *Journal of Vision*, *6*(12):9, 1431–1450, doi:10.1167/6.12.9. [PubMed] [Article]

Kit, D., Katz, L., Sullivan, B. T., Snyder, K., Ballard, D. H., & Hayhoe, M. M. (2014). Eye movements, visual search and scene memory, in an immersive virtual environment. *PLoS One*, *9*, 1–18, doi:10.1371/journal.pone.0094362.

Kunar, M. A., Flusberg, S., & Wolfe, J. M. (2008). The role of memory and restricted context in repeated search. *Perception & Psychophysics*, *70*, 1117–1129, doi:10.3758/PP.

Land, M. F. (2004). The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental Brain Research*, *159*, 151–160, doi:10.1007/s00221-004-1951-9.

Mack, S. C., & Eckstein, M. P. (2011). Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision*, *11*(9):9, 1–16, doi:10.1167/11.9.9. [PubMed] [Article]

Mou, W., McNamara, T. P., Valiquette, C. M., & Rump, B. (2004). Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 142–157.

Neider, M. B., & Zelinsky, G. J. (2008). Exploring set size effects in scenes: Identifying the objects of search. *Visual Cognition*, *16*, 1–10, doi:10.1080/13506280701381691.

Olejarczyk, J. H., Luke, S. G., & Henderson, J. M. (2014). Incidental memory for parts of scenes from eye movements. *Visual Cognition*, *22*, 975–995, doi:10.1080/13506285.2014.941433.

Olson, I. R., & Chun, M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, *9*, 273–302, doi:10.1080/1350628004200016.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442, doi:10.1163/156856897X00366.

Pinheiro, J. C., & Bates, D. M. (2000). *Statistics and computing: Mixed-effects models in S and S-Plus. Statistics and computing*. Berlin: Springer-Verlag.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision*, *7*(14):16, 1–20, doi:10.1167/7.14.16. [PubMed] [Article]

Solman, G. J. F., & Kingstone, A. (2014). Balancing energetic and cognitive resources: Memory use during search depends on the orienting effector. *Cognition*, *132*, 443–454, doi:10.1016/j.cognition.2014.05.005.

Solman, G. J. F., & Smilek, D. (2012). Memory benefits during visual search depend on difficulty. *Journal of Cognitive Psychology*, *24*, 689–702, doi:10.1080/20445911.2012.682053.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*(5):5, 1–23, doi:10.1167/11.5.5. [PubMed] [Article]

Tatler, B. W., & Tatler, S. L. (2013). The influence of instructions on object memory in a real-world setting. *Journal of Vision*, *13*(2):5, 1–13, doi:10.1167/13.2.5. [PubMed] [Article]

Võ, M. L.-H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision*, *10*(3):14, 1–13, doi:10.1167/10.3.14. [PubMed] [Article]

Võ, M. L.-H., & Wolfe, J. M. (2012). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 23–41, doi:10.1037/a0024147.

Võ, M. L.-H., & Wolfe, J. M. (2013). The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition*, *126*, 198–212.

Waller, D., & Hodgson, E. (2006). Transient and enduring spatial representations under disorientation and self-rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 867, doi:10.1037/0278-7393.32.4.867.

Williams, C. C., Henderson, J. M., & Zacks, R. T. (2005). Incidental visual memory for targets and distractors in visual search. *Perception & Psychophysics*, *67*, 816–827, doi:10.1016/j.biotechadv.2011.08.021.Secreted.

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception & Psychophysics*, *73*, 1650–1671, doi:10.3758/s13414-011-0153-3.

Appendix A

Bootstrapping the parameters for the exponential model fit to the learning curve data in Figure 4A and 4B.

The function we fit takes the form of $y = ix^s$ Here, y can either be number of search fixations or search time, i is the intercept and s is the slope of the function, and x is search trial. i and s are the outputs from nonlinear mixed-effect modeling (NLME; Pinheiro & Bates, 2000). The experiment (2D vs. 3D) was the factor specified as the fixed effect, and the difference between subjects was the random effect included. For the number of search fixations in the correct room (Figure 4A, top), the intercept of fit produced by NLME is 5.21 in 3D and 10.08 in 2D; the slope is −0.26 in both 3D and 2D. The bootstrapped difference between the distributions of intercepts in 3D and that in 2D is significant ($M = -4.8$, $SE = 0.53$, $p < 0.001$), whereas the bootstrapped difference of distributions of slopes is not different ($M = 0$, $SE = 0.01$, $p = 0.48$). For search time in the correct room (Figure 4A, bottom), the intercept obtained is 2.81 in both 2D and 3D; the slope is −0.3 in 3D and −0.38 in 2D. The bootstrapped difference of intercepts is not significant ($M = 0.09$, $SE = 0.22$, $p = 0.34$), as well as for slopes ($M = 0.05$, $SE = 0.05$, $p = 0.14$). For number of fixations in the incorrect room (Figure 4B, top), the intercept obtained is 24.58 in 3D and 11.76 in 2D; the slope is −0.8 in 3D and −0.44 in 2D. The bootstrapped difference between the distributions of intercepts in 3D and that in 2D is significant ($M = 10.88$, $SE = 5.7$, $p = 0.03$), and the bootstrapped difference of distributions of slopes is also different ($M = -0.29$, $SE = 0.17$, $p = 0.04$). For search time in the incorrect room (Figure 4B, bottom), the intercept obtained is 13.37 in 3D and 3.27 in 2D; the slope is −0.78 in 3D and −0.52 in 2D. The bootstrapped difference of intercepts is significant ($M = 9.36$, $SE = 2.5$, $p < 0.001$), as well as for slopes ($M = -0.14$, $SE = 0.17$, $p < 0.001$).

Appendix B

Bootstrapping the difference between the search performances of novel search for geometric objects, repeated search for geometric objects, search for nearby contextual objects, and search for distant contextual objects (data shown in Figure 7).

We suspected that early search experience for geometric objects might aid subsequent search for other items in the virtual apartment as a result of memory representations of the environment being developed through the experience. We thus compared search performance for novel searches for geometric objects (Trials 1–8), repeated searches for geometric objects (Trials 9–24), search for nearby contextual object (eight trials within Trials 25–40), and distant contextual objects (eight trials within Trials 25–40), in both the correct room and the incorrect room. Because those four groups of samples are not independent (one covariate is time but the change is not linear: Repeated geometric objects were sought for after novel search trials, the timing of search for nearby and distant contextual objects was mixed within participants, and the order of trials were randomized between participants), we did not use a one-way ANOVA to analyze our data first. Instead, we calculated the obtained F value from our data, bootstrapped the F distribution from 5,000 repetitions of resampling from the data, and calculated the p value by evaluating the percentage of area under the bootstrapped F distribution that is larger than the obtained F value. The results generated showed the same trend as the one-way ANOVA we

later conducted; thus, we report the ANOVA results below. There were significant differences between both number of search fixations and search time for novel geometric search, repeated geometric search, nearby contextual search, and distant contextual search in the incorrect room and the correct room, in both 3D—(fixation/correct room): $F(3, 68) = 4.11$, $p = 0.01$, fixation/incorrect room: $F(3, 68) = 12.56$, $p < 0.001$, time/correct room: $F(3, 68) = 2.96$, $p = 0.04$, time/incorrect room: $F(3, 68) = 8.3$, $p < 0.001$—and 2D—fixation/correct room: $F(3, 76) = 30.35$, $p < 0.001$, fixation/incorrect room: $F(3, 76) = 16.72$, $p < 0.001$, time/correct room: $F(3, 76) = 32.91$, $p < 0.001$, time/incorrect room: $F(3, 76) = 15.54$, $p < 0.001$.

Then we bootstrapped the pairwise differences of the four conditions by resampling the data for 5,000 repetitions for each pair, acquiring the sampling distributions of the means of the differences and derived p values from the means and standard errors of those distributions. The results showed that in both rooms in 3D (see Figure 5A), the search for nearby contextual objects was not as efficient as the repeated search for geometric objects (fixation/correct room: $M = 1.54$, $SE = 0.77$, $p = 0.02$; fixation/incorrect room: $M = 4.37$, $SE = 1.42$, $p = 0.001$; time/correct room: $M = 2.44$, $SE = 0.86$, $p = 0.002$; time/incorrect room: $M = 3.13$, $SE = 1.12$, $p = 0.003$) and was better than novel search for geometric objects only in terms of number of search fixations in the incorrect room ($M = -4.47$, $SE = 1.97$, $p = 0.01$) but not in other parameters (fixation/correct room: $M = 0.31$, $SE = 0.82$, $p = 0.35$; time/correct room: $M = 0.72$, $SE = 0.97$, $p = 0.23$; time/incorrect room: $M = -1.4$, $SE = 1.31$, $p = 0.14$). Search efficiency for distant contextual objects in 3D also did not benefit from previous search experience, because it is no more efficient than novel search for geometric objects, except for number of fixations in the incorrect room (fixation/correct room: $M = 1.97$, $SE = 1.1$, $p = 0.03$; fixation/incorrect room: $M = -3.31$, $SE = 1.67$, $p = 0.02$; time/correct room: $M = 1.39$, $SE = 0.99$, $p = 0.08$; time/incorrect room: $M = -1.58$, $SE = 0.91$, $p = 0.04$); it is also less efficient than repeated search for geometric objects (fixation/correct room: $M = 3.23$, $SE = 1.06$, $p = 0.001$; fixation/incorrect room: $M = 5.48$, $SE = 1.07$, $p < 0.001$; time/correct room: $M = 3.11$, $SE = 0.89$, $p < 0.001$; time/incorrect room: $M = 2.95$, $SE = 0.62$, $p < 0.001$). In 2D, the search for neither nearby nor distant contextual objects benefits from early searches for geometric objects.